First Insigth in the conserved periplasmic domain of the *Escherichia coli* Pal protein

The periplasmic peptidoglycan-associated lipoprotein (Pal) from *Escherichia coli* is part of the Tol-PAL multiprotein complex used by group A colicins to penetrate and kill cells. Out of the seven genes corresponding to the Tol-PAL operon, five genes were submitted to a number of experiments, including knock-out and mutations demonstrating their implication in the bacterial envelope integrity. These genes are specific of the Gram-negative bacteria and conserved in many of them. The structure of the TolB protein as well as the C-terminal periplasmique domain of the TolA protein were solved earlier and permitted to generate some hypothesis on the molecular function of these proteins. In order to better understand the protein interactions involved in the Tol/PAL system we initiated the structure determination of another component of this system, the PAL protein

The PAL protein is a 152 amino-acid long protein. We solved the crystal structure of the C-terminal 109 amino acid fragment of the Pal protein using multiwavelength anomalous Dispersion (MAD) method and refined the structure to 1.93 Å. The Pal protein forms a crystallographic dimer in the crystal. The Pal structure is a four stranded β–sheet floor flanked by three α-helices.

Data collection: Crystal were collected in a Hampton Research 0.5 mm^3 loop, flash frozen to 105K in a cold nitrogen gas stream and subjected to X-ray diffraction.

The MAD data sets were collected on a MAR CCD detector at the ESRF radiation synchrotron facility (ID29) on a synchrotron beamline (ESRF, Grenoble-France). The peak and inflection wavelength of the Se absorption edge as well as the remote were selected based on the fluorescence spectrum of the Pal crystal (Table 1). The native data set was collected on a MAR Research imaging-plate detector on a Rigaku RU-200 rotating-anode generator running at 40kV and 80mA with a copper target at a wavelength of 1.5418Å. Both the native and MAD data sets were processed using the *MOSFLM* package; and the *SCALA* program from the *CCP4* package was used for the scaling and data reduction.

Structure Determination and refinement: Initial positions of the selenium sites were obtained using the SOLVE with subsequent refinement of the sites and phase calculation with

the autoSHARP programs. A single solution was found by SOLVE with 2 sites and a mean figure of merit of 0.33 for all the data between 20-2.5 Å. The atom positions were used in autoSHARP and the phase obtained (correlation 0.14) were improved using solvent-flattening and histogram-matching techniques as implemented in the autoSHARP program. The final correlation is of 0.72. The electron-density maps were used to construct the main chain of the molecules using TURBO-FRODO. Preliminary refinement was performed using CNS between 20 and 2.5 Å and we subsequently used the native data set (18 to 1.93 Å) for final refinement. The quality of the structure was checked using PROCHECK, as a result 95 residues (99%) of the Pal molecule occupy the most favored areas of the Ramachandran plot, and 1 is in the generously allowed region (first residue of the Pal structure). The conformations of most residues were well defined except for the 2 N-terminal residues.

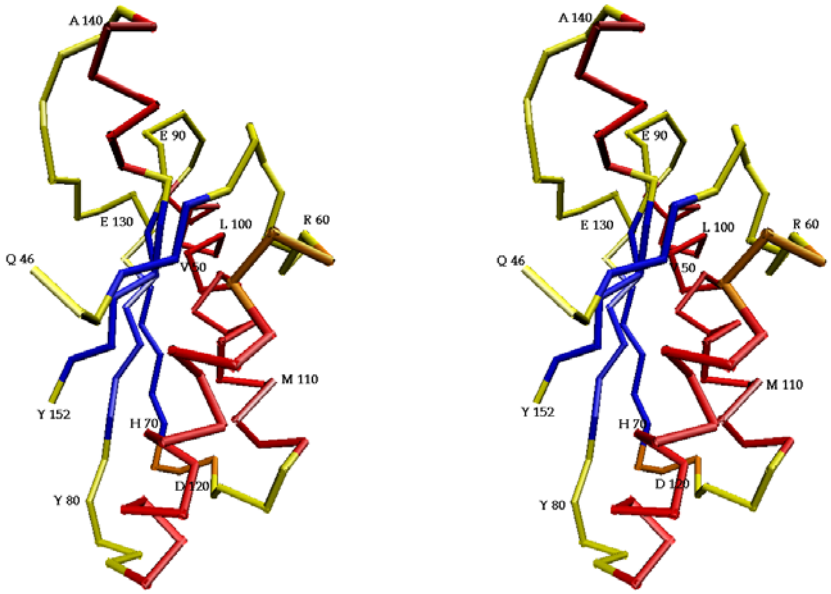

Table 1 X-ray data collection and refinement statistics

Data collection				
Data set	Native		MAD	
		λ1	λ2	λ3
Beam line	Local		ESRF/ID29	
Wavelength (Å)	1.5418	0.979213	0.979527	0.93928
Space group	$I4_122$		$I4_122$	
Unit cell dimensions(Å)	a=b=88.58		a=b=89.17	
	c=68.03		c=68.98	
Resolution range (Å)	18-1.93	34.5-2.5	34-2.5	34-2.5
Observations	118335	312240	312240	312240
Unique reflections	10342	9373	9383	9422
Multiplicity[1]	5.8 (5.0)	4.0 (4.0)	4.1 (4.1)	4.1 (4.1)
Completeness[1]	99.1 (92.4)	95.9 (97.5)	96.0 (97.5)	96.3 (97.5)
Anomalous Completeness[1]		75.6 (33.3)	75.8 (33.4)	76.1 (33.4)
$\langle I / \sigma I \rangle$[1,2]	7.5 (3.0)	10.3 (6.7)	8.2 (6.1)	7.7 (3.4)
R_{sym} (%)[1,3]	5.2 (25)	4.3 (10.6)	5.2 (11.5)	5.5 (21.8)
Refinement				
R_{cryst} (%)[4]	20.0			
R_{free} (%)	23.4			
Δ_{bond} (Å)	0.018			
Δ_{angle} (°)	1.7			
Non-hydrogen atoms	2			
Protein	860			
water	86			
Mean Bfactor (Å^2)	30.4			

[1] values in parentheses are for the highest resolution shell.

[2] $\langle I / \sigma I \rangle$, is the mean signal to noise ratio, where I is the integrated intensity of a measured reflection and σ is the estimated error in the measurement.

[3] $R_{sym} = \sum_h \sum_i \left| I_{h,i} - \langle I_h \rangle \right| / \sum_h \sum_i \left| I_{h,i} \right|$, where I is the integrated intensity of reflection h having i observations and $\langle I_h \rangle$ is the mean recorded intensity of reflection h over multiple recording.

[4] $R_{cryst} = \sum \left| |F_o| - |F_c| \right| / \sum |F_o|$, where F_o are observed and F_c calculated structure factor amplitudes. R_{free} is calculated from a randomly chosen 9.9% of reflections.